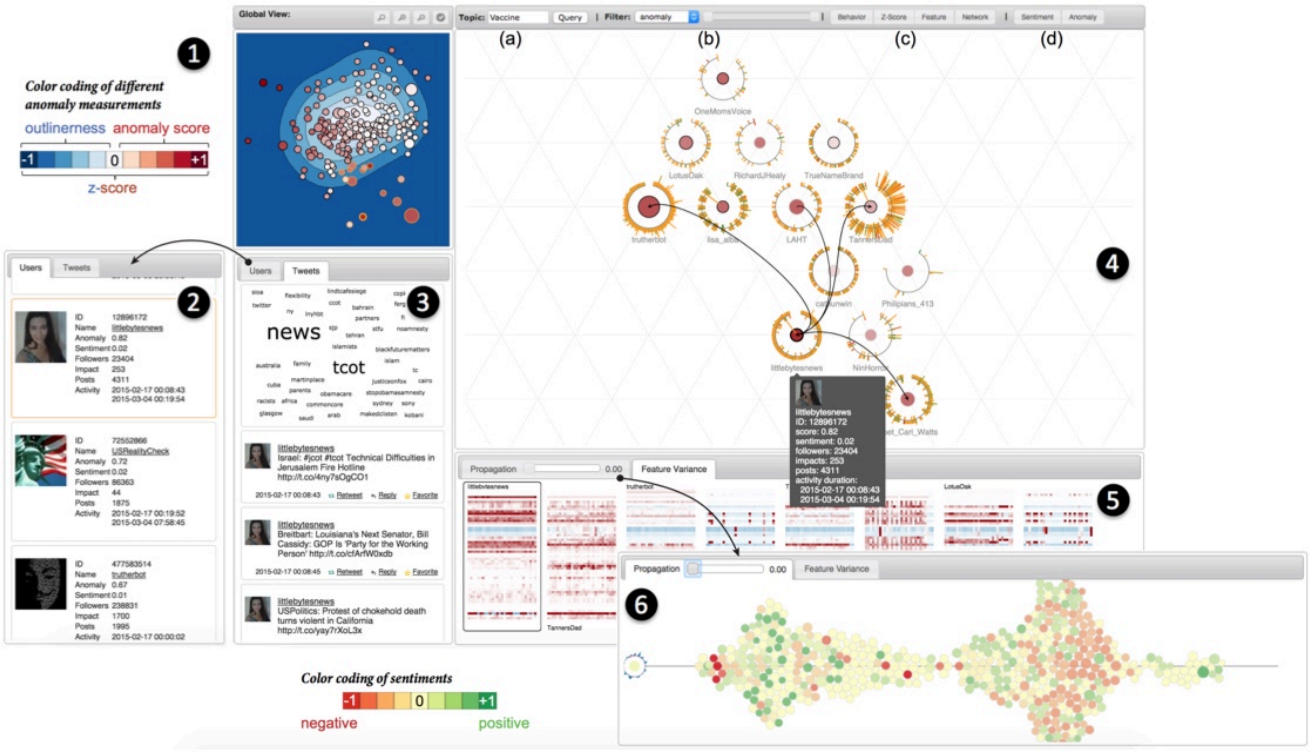
visual analysis of anomalous user behaviors in online communication systems



发表

TVCG 2015

Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, Ching-Yung Lin

摘要

detection of users with anomalous communication behaviors

1. Twitter data
2. email records

挑战

1. display and capture the rich contexts of a communication process through a simple and integrated visual design to facilitate efficient visual comparison

通过简单完成的可是设计去对富文本的交流过程进行有效的可视比较

2. capture those patterns are important for revealing the insight of a user's behavior
    - temporal patterns
    - content patterns
    - activity patterns 几种pattern的获取
3. a standard approach or common understanding of the underlying structures of a typical communication process is lacking
   典型交流过程的基本结构的标准未知，如何去构建可以应用于不同communication systems的可视化

背景

- a novel visual analysis system for **detecting, summarizing, interpreting**, and **comparing** anomalous user behaviors archived in various types of communication data
- an unsupervised learning model, **TLOF**

贡献

1. system
2. visualization
    - new glyph designs
        - activity glyph
        - z-glyph
        - relation glyph
    - layout
        - based on their similarities in a triangle grid
3. evaluation
    - a bot detection challenge on Twitter
    - case study based on an email dataset
    - visual patterns of the anomalous user behaviors

相关工作

1. Anomaly Detection

2. Visual Analysis of User Behaviors
3. Visual Summarization of Activities

需求

和专家讨论得到

1. Feature Selection
2. Anomaly Detection in Contexts.
3. Ranking Threats
    - Semantics
4. Learn from User Feedback

系统

1. the data collection module
2. the preprocessing module
3. the analysis module
4. the visualization module.

用户行为

方法

Time-Adaptive Local Outliner Factor. **TOLF**

- 无监督学习
- 考虑了时序
- 分数而不是label
- 欧氏距离易于被可视化解释

feature

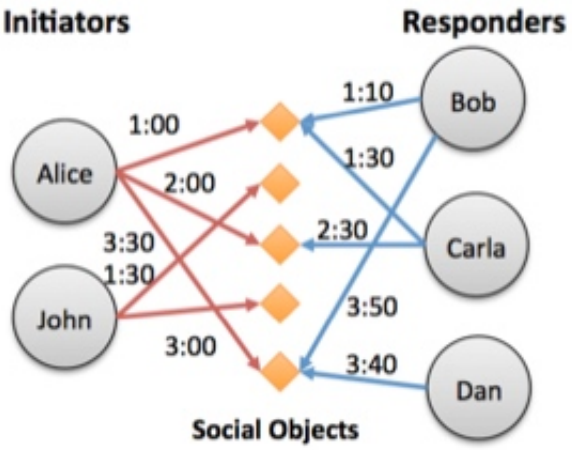
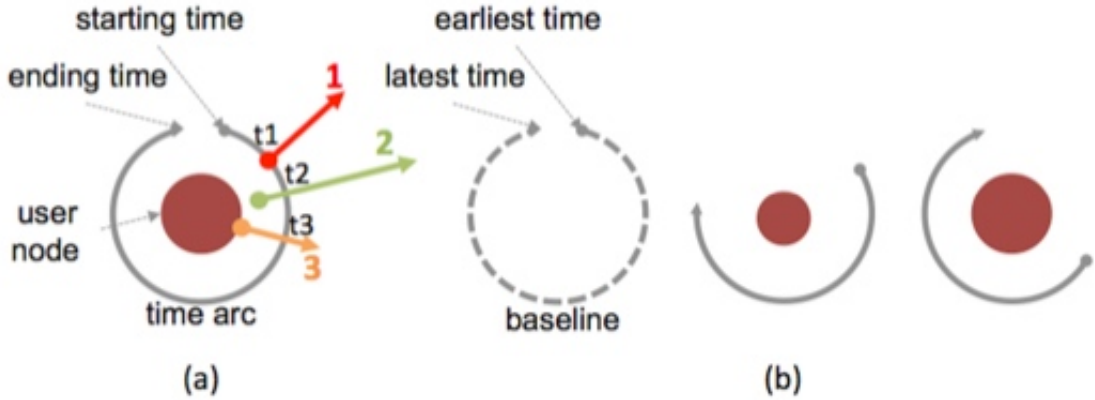


Fig. 3. Data model for social communication.

1. Behavior Features.
   角色分类
2. Content Features
   关键词转变
3. Interaction Features
4. Temporal Features
5. Network Features
   ego-centric measurements
6. User Profile Features
   用户信息改变

设计

1. size 编码重要性
2. color 编码情绪或者异常分数
3. users are visualized as glyphs
   - Behavior Glyph.



- Z-Glyph
- Relation Glyph.

4. layout
   - **mesh**网格有助于快速线性布局，消除节点重复
   - **triangle mesh**三角网格，preserve topologies of any surface or shape.捕捉到glyphs的拓扑上的相似度

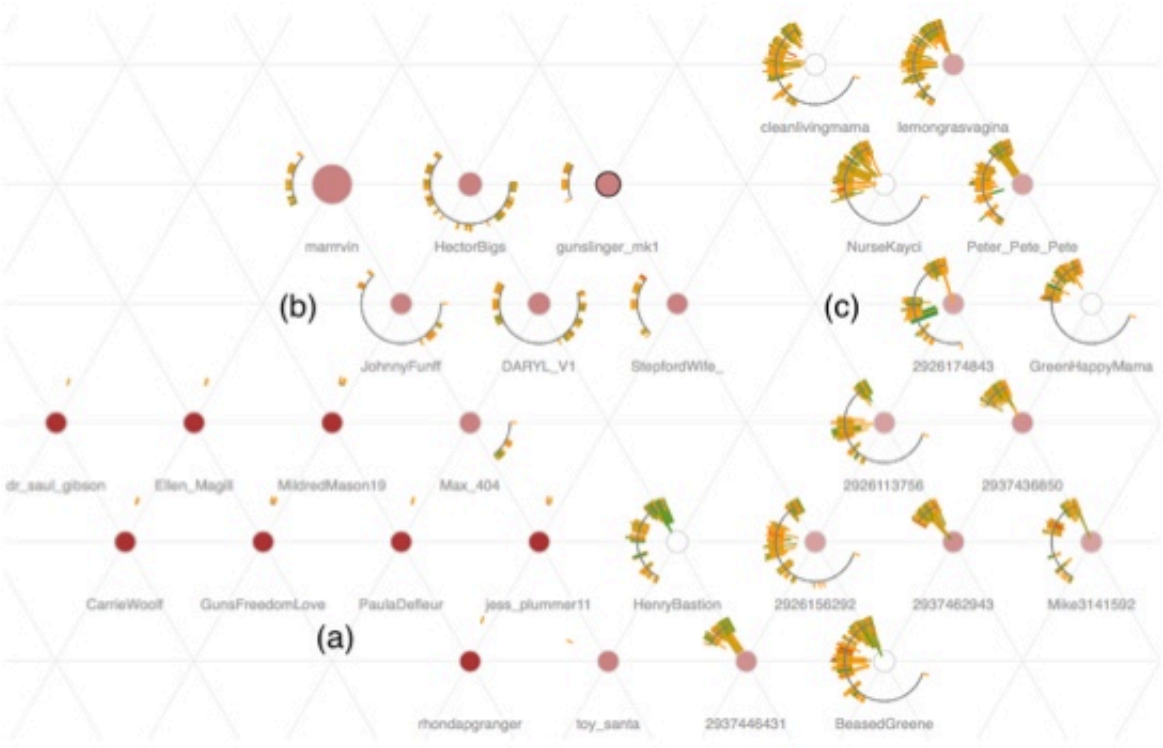


Fig. 10. The behaviors of all the social bots

交互功能

- query
- Filter
- Highlight
- Inspection
- Switch Contexts
- Data Index
- Zoom and Pan